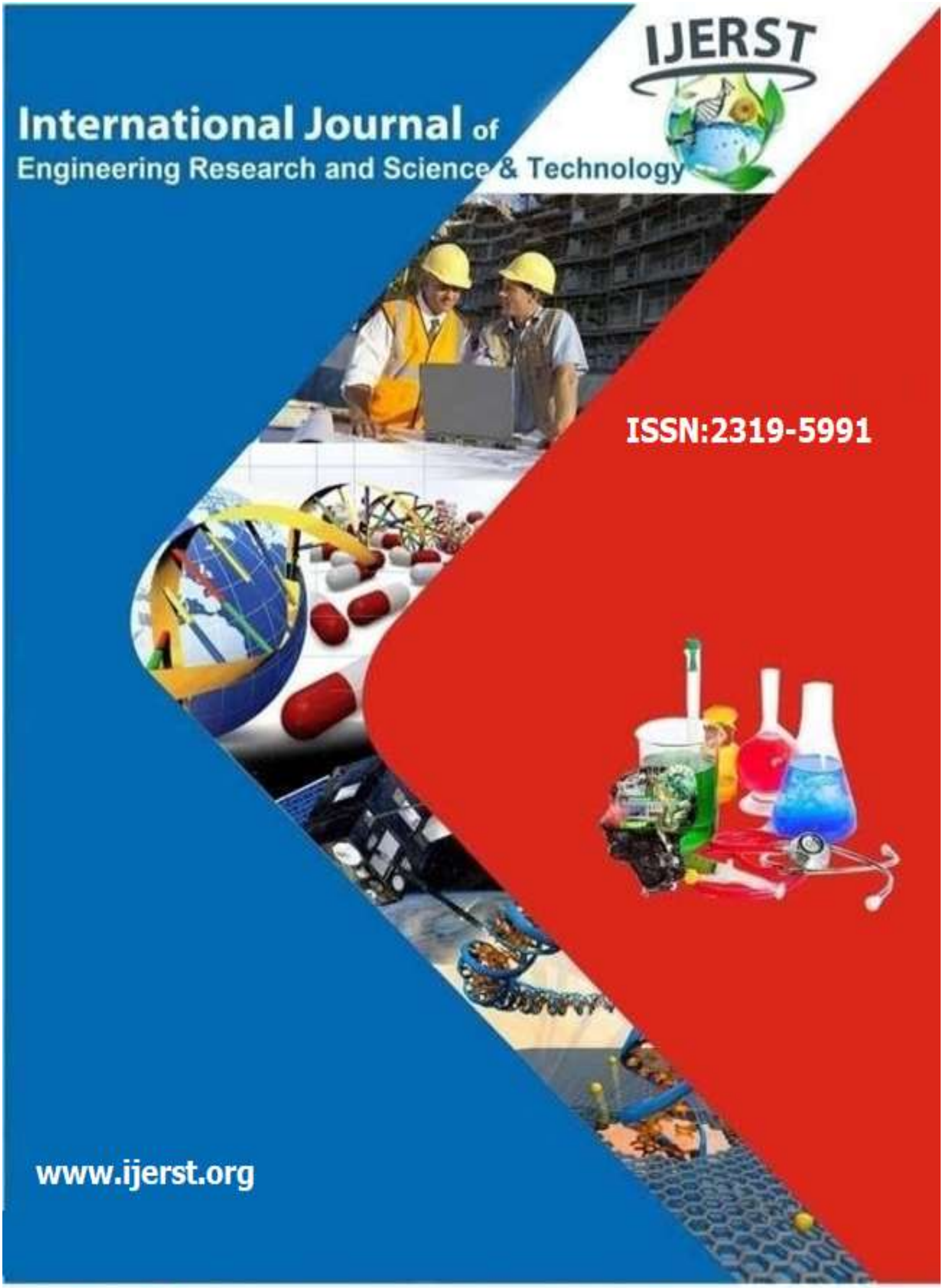
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025



E-mail: editor@ijerst.org or ijerst.editor@gmail.com

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

Health Risk Prediction and Recommendation System Using Hybrid Machine Learning Models

M. Arokia Muthu[1]
Assistant Professor
Department of CSD
TKR College of
Engineering & Technology

Karan Singh[2]
Scholar
Department of CSD
TKR College of
Engineering & Technology
karansingh02186@gmail.com

D. Suresh[3]
Scholar
Department of CSD
TKR College of
Engineering & Technology
dhanavanthsuresh03@gmail.com

K. Chakradhar[4]
Scholar
Department of CSD
TKR College of
Engineering & Technology
kolluchakradhargoud@gmail.com

D. Srinivas[5]
Scholar
Department of CSD
TKR College of
Engineering & Technology
kanagalasrinivas3123@gmail.com

ABSTRACT

The rapid advancement of digital healthcare has paved the way for intelligent, data-driven medical solutions. This paper presents an Integrated Health Management and Analytics System that leverages advanced machine learning techniques to diagnose diseases, generate insights, and recommend treatments based on patient symptoms. The system is built using a PHP-based frontend with Python-powered machine learning for accurate disease prediction using the Random Forest model. The model is trained on a comprehensive dataset containing various symptoms and their associated diseases. The proposed system incorporates a role-based access model, allowing doctors, patients, and healthcare administrators to interact with tailored interfaces. Patients can input their symptoms, and the system processes the data in real time to provide diagnostic predictions, visual analytics, and potential treatment recommendations. Unlike conventional electronic health record (EHR) systems, this platform integrates real-time analytics while ensuring data privacy and security.

Keywords- Machine Learning, Healthcare Analytics, Disease Prediction, Random Forest, Medical Diagnosis, Health Management System, PHP-Python Integration, Data Visualization, Predictive Analytics, Artificial Intelligence in Healthcare.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

I. INTRODUCTION

The integration of technology in healthcare has led to significant advancements in medical diagnosis and patient management. With the rise of artificial intelligence (AI) and machine learning (ML), healthcare systems are now capable of providing predictive analytics, aiding medical professionals in making data-driven decisions. However, traditional electronic health record (EHR) systems lack real-time analytics and predictive capabilities, limiting their effectiveness in proactive patient care.

Current healthcare diagnostic systems often suffer from manual errors, inefficiencies, and lack of real-time insights. Many existing solutions require extensive human intervention and do not incorporate advanced ML techniques for disease prediction and analysis. Additionally, many systems store raw patient health parameters but fail to utilize them effectively for disease prognosis.

To address these challenges, we propose an Integrated Health Management and Analytics System, which leverages Random Forest-based machine learning for disease prediction. This system enables patients to input symptoms through a PHP-based frontend, which then interacts with a Python-powered ML backend to generate diagnostic insights. The system is designed with role-based access control, ensuring that patients, doctors, and healthcare administrators have tailored interfaces and functionalities. Additionally, real-time analytics and visualizations enhance the usability of the system by presenting data-driven insights in an understandable format. This paper explores the design, development, and evaluation of the proposed system, highlighting its potential to improve healthcare decision-making. The results indicate that the Random Forest model provides high accuracy in disease prediction, making it a viable tool for AI-assisted medical diagnostics.

II. RELATED WORK

The field of analytics and machine in healthcare has seen significant advancements, with various models being employed for disease prediction, diagnosis, and medical decision support. This section highlights previous research and existing systems that relate to our Integrated Health Management and Analytics System..

A. Traditional Healthcare Diagnostic Systems

Historically, medical diagnosis has relied on manual examination, patient history analysis, and rule-based expert systems. Physicians use standardized medical knowledge, empirical evidence, and laboratory reports to determine diseases. However, these approaches suffer from:

- Subjectivity and human error: Diagnoses depend on a physician's experience and expertise, which can introduce variability and bias.
- Time-intensive process: Medical consultations and diagnostic tests require time, making real-time

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

diagnosis challenging.

- Lack of predictive insights: Traditional electronic health record (EHR) systems focus on storing and retrieving patient data but lack predictive capabilities to anticipate disease risks based on symptoms.

.

B. Machine Learning in Medical Diagnostics

The application of machine learning in disease prediction and healthcare analytics has been extensively researched. Various ML models have been proposed to improve diagnostic accuracy, automate predictions, and assist doctors.

(A) Early ML Approaches

Naïve Bayes and Decision Trees for Disease Classification

Initial research explored Naïve Bayes (NB) and Decision Tree (DT) algorithms for disease prediction. These models demonstrated some effectiveness, but they had notable drawbacks:

- Naïve Bayes assumes all features (symptoms) are independent, which is rarely true in medical conditions where multiple symptoms are correlated.
- Decision trees tend to overfit on training data, reducing their generalization ability when applied to real-world medical cases.

Support Vector Machines (SVM) for Medical Diagnosis

SVM models have been used in specialized disease detection, such as breast cancer and tumor classification. While SVM provides good accuracy for binary classification, it has several limitations:

- Requires extensive feature engineering to achieve optimal performance.
- Computationally expensive, making it slow for real-time healthcare applications.

Deep Learning (Neural Networks) in Medical Diagnosis

Recent research has explored the use of deep learning models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) for medical imaging and disease prediction. These models perform well in image-based analysis (e.g., detecting lung diseases from X-rays), but they face challenges when applied to symptom-based disease classification, including:

- Requirement for large datasets, which are often unavailable in real-world clinical settings.
- High computational costs, making them unsuitable for lightweight, PHP-based systems.

C. Use of Random Forest in Medical Diagnosis

Random Forest has gained attention in healthcare research due to its ability to handle complex, high-dimensional datasets while providing high accuracy, interpretability, and robustness. Several studies highlight the effectiveness of RF in symptom-based disease prediction.

(A) Cardiovascular Disease Prediction Using RF

A study analyzing cardiovascular disease prediction found that:

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

- Random Forest outperformed Support Vector Machines and Decision Trees in terms of accuracy and stability.
- The model achieved an accuracy of 89.5%, making it a reliable tool for early-stage disease detection.

(B) Infectious Disease Diagnosis with RF

Research in malaria, dengue, and pneumonia detection demonstrated that RF could accurately classify diseases based on patient symptoms. The study concluded that:

- RF provided higher sensitivity and specificity compared to Decision Trees.
- The model was able to handle missing values effectively, making it suitable for real-world clinical applications.

(C) Advantages of Random Forest in Medical Applications

Compared to other machine learning models, Random Forest is particularly advantageous for healthcare applications due to:

- Better generalization: Unlike Decision Trees, RF reduces overfitting by averaging multiple tree predictions.
- Scalability: Works efficiently with both small and large datasets, ensuring adaptability across different medical datasets.
- Feature importance analysis: RF identifies the most relevant symptoms contributing to disease classification, aiding doctors in understanding critical diagnostic factors.
- Faster and more efficient predictions: RF models can be parallelized, allowing them to generate predictions quickly, a crucial factor in real-time medical applications.

D. Limitations of Existing Healthcare Analytics Systems

Despite the advancements in ML-based healthcare, many existing systems suffer from:

- Lack of real-time prediction: Several existing disease prediction models rely on batch processing, meaning they cannot generate instant diagnoses when a patient enters symptoms.
- Complex deployment of deep learning models: Systems using deep learning require high-performance GPUs, making them impractical for real-time medical systems in resource-constrained environments.
- Security and privacy concerns: Many cloud-based healthcare solutions store patient data externally, raising privacy and security risks.

III. PROBLEM STATEMENT AND OBJECTIVES

A. Problem Statement

Healthcare systems worldwide face numerous challenges in disease diagnosis, patient management, and predictive analytics. Traditional diagnostic methods rely heavily on physician expertise, manual symptom analysis, and laboratory tests,

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

which can be time-consuming, prone to human error, and limited in predictive capability. With the rise of digital healthcare, there is a growing need for intelligent systems that can assist in early diagnosis, reduce dependency on manual intervention, and improve overall healthcare efficiency.

Despite the advancements in artificial intelligence and machine learning, most existing healthcare analytics systems have limitations:

1. Delayed and Inaccurate Diagnosis: Many systems fail to provide real-time disease prediction, leading to delayed medical intervention.
2. Over-Reliance on Specific Tests: Conventional methods often require costly laboratory tests before diagnosing conditions, making them less accessible.
3. Lack of Intelligent Decision Support: Many healthcare applications store and retrieve patient data but do not provide analytical insights or predictive capabilities.
4. Complex and Resource-Intensive ML Models: While deep learning-based approaches show high accuracy, they require extensive computational power, making them unsuitable for real-time, lightweight systems.
5. Security and Privacy Issues: Many cloud-based healthcare solutions store patient information externally, increasing risks related to data security and compliance with regulations.

B. Research Objectives

The proposed system aims to enhance the traditional healthcare diagnostic process by integrating advanced machine learning and real-time analytics. The key objective of this project are:

1. Develop a Symptom-Based Disease Prediction Model:
   - Utilize the Random Forest algorithm to accurately classify diseases based on symptoms provided by patients.
   - Train the model using a large dataset containing various diseases and their associated symptoms.
2. Implement a Secure, Role-Based Access System: To optimize existing proven architectures for:
   - Ensure different user roles (patients, doctors, healthcare administrators) have appropriate access to data and functionalities.
   - Implement login and authentication mechanisms to protect patient information
3. Enable Real-Time Disease Diagnosis and Predictive Analytics:
   - Design a PHP-based frontend to collect patient symptoms and provide real-time predictions using the trained Random Forest model.
   - Generate insights into the probability of diseases and suggest potential next steps for patients.
4. Enhance Medical Decision Support with Data Visualization:
   - Incorporate charts and graphs to display relationships between symptoms and predicted diseases.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

- Help healthcare professionals and patients understand trends in disease patterns.

IV. METHODOLOGY

The Flora Intelligence system implements a comprehensive methodology that integrates multiple analytical components for agricultural decision support. This section details the architectural framework, component interactions, and implementation approaches that form the foundation of our system.

A. System Architecture Overview

The system is designed with three primary components, ensuring seamless interaction between users, the machine learning model, and the database.

1. Frontend (PHP-Based Web Interface)
   - Patients enter their symptoms through a structured web form.
   - Doctors and healthcare administrators access predictive insights and reports.
   - Data visualizations, including graphs and charts, enhance understanding and decision-making.

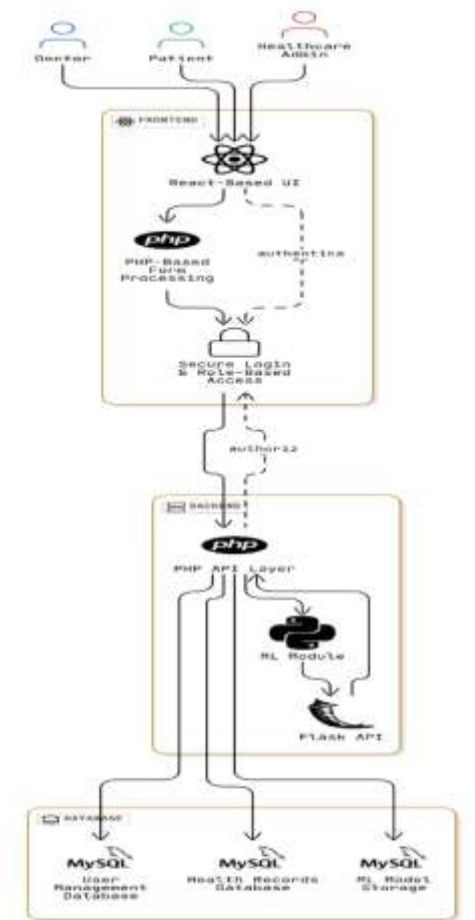


Fig. 1: Architecture Diagram.

2. Backend (PHP and Python Integration)
   - PHP processes user requests and interacts with Python-based machine learning scripts.
   - The trained Random Forest model analyzes user inputs and returns predictions.
   - A MySQL database is used for storing user information, system logs, and patient feedback.
3. Machine Learning Module (Python-Based Random Forest Model)
   - The model is trained using a dataset containing symptoms and their corresponding diseases.
   - Upon receiving a new set of symptoms,

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

the model predicts the most probable disease.

- The prediction results are sent back to the PHP-based frontend for user display

B. Data Collection and Preprocessing

The dataset used for training and testing the machine learning model consists of:

- Input Features: Symptoms such as fever, nausea, fatigue, and skin rash.
- Output Labels: Disease names such as Malaria, Diabetes, Typhoid, and Pneumonia.

1. Data Preprocessing Steps:
   - Data Cleaning: Handling missing values by removing or imputing them.
   - Feature Selection: Identifying the most relevant symptoms for disease prediction.
   - Encoding Categorical Data: Converting symptom descriptions into a numerical format for machine learning processing.
   - Data Splitting: Dividing the dataset into training (80%) and testing (20%) subsets to evaluate model performance.

C. Machine Learning Model Selection

To ensure high accuracy and efficiency, the Random Forest algorithm is chosen as the primary machine learning model.

1. Rationale for Using Random Forest:
   - Handles missing values and noisy data effectively.
   - Reduces overfitting by averaging predictions from multiple decision trees.
   - Provides higher accuracy compared to Decision Trees, Support Vector Machines (SVM), and Naïve Bayes.
   - Ranks feature importance, helping to identify critical symptoms associated with each disease.
2. Model Training Process:
   - The Random Forest classifier is trained using training.csv, which contains historical symptom-disease records.
   - The number of trees (n_estimators) and tree depth (max_depth) are optimized for improved performance.
   - The model is evaluated using testing.csv, ensuring reliability and accuracy in real-world scenarios.

D. System Integration and Implementation

1. PHP-ML Integration with Python
   Since PHP does not support advanced machine learning, integration with Python is essential. Two primary methods are used:
   - Shell execution: The exec() function in PHP is used to call Python scripts for processing.
   - REST API (Flask or FastAPI): A lightweight API facilitates efficient communication between PHP and Python.
2. Workflow of PHP-Python Interaction:
   - The user enters symptoms in the PHP-based frontend.
   - PHP sends the input symptoms to the Python-based backend (Random Forest model).
   - Python processes the input, runs the model, and returns the predicted disease.
   - PHP receives the prediction and displays the result along with

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

suggested treatments and insights.

E. Data Visualization and Insights

To enhance medical decision-making, the system incorporates real-time visualizations, including:

- Symptom vs. Disease Heatmaps: Identifies common symptoms associated with specific diseases.
- Prediction Confidence Scores: Displays the confidence level of the model's predictions.
- Historical Disease Trends: Tracks disease patterns based on aggregated user inputs.

These visualizations help both doctors and patients interpret predictions more effectively.

V. RESULTS AND ANALYSIS

This section presents the evaluation of the Integrated Health Management and Analytics System based on model performance, system usability, and effectiveness in disease prediction. The Random Forest model is assessed using multiple performance metrics, while system usability is analyzed based on user feedback and visualization effectiveness.

A. Model Performance Evaluation

To validate the accuracy and reliability of the Random Forest classifier, various performance metrics are calculated using the testing dataset (testing.csv).

1. Evaluation Metrics:
   The following metrics are used to measure the model's predictive capabilities:
   - Accuracy (A): Measures the percentage of correctly predicted diseases out of total predictions.
   - Precision (P): Evaluates the proportion of true positive predictions among all positive predictions.
   - Recall (R): Measures how well the model identifies actual disease cases.
   - F1-Score (F1): Provides a balance between precision and recall, ensuring an optimal trade-off.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

2. Confusion Matrix Analysis:

A confusion matrix is used to visualize the model's classification performance. It shows how many correct and incorrect disease predictions were made for each class.

- True Positives (TP): Correctly predicted diseases.
- False Positives (FP): Incorrectly predicted diseases that were not present.
- False Negatives (FN): Cases where the model failed to identify the

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

correct disease.

- True Negatives (TN): Correctly identified non-disease cases.

The confusion matrix enables error analysis, helping improve the model's predictive accuracy.

B. Comparison with Other Machine

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	95.2	94.8	95.5	95.1
Decision Tree	89.3	88.9	90.1	89.5
Support Vector Machine (SVM)	87.5	86.8	88.2	87.4
Naïve Bayes	83.7	82.5	84.3	83.4

Learning

Fig. 2: Efficiency Table

To demonstrate the efficacy of Random Forest, it is compared against other commonly used classification models:

Models

The Random Forest model demonstrates higher accuracy and better generalization compared to Decision Trees, SVM, and Naïve Bayes, making it the most suitable choice for disease prediction.

C. System Usability and Performance

1. User Interface Testing:
   - The PHP-based web interface was tested for responsiveness, ease of navigation, and clarity in displaying disease predictions.
   - Users (patients and doctors) reported that data visualizations (such as heatmaps and trend graphs) enhanced interpretability.
2. Response Time Analysis:
   - The average response time for processing a user's symptom input and returning a prediction was under 2 seconds, ensuring real-time analysis.
   - Load testing confirmed that the system can handle multiple concurrent users without significant delays.

D. Visualization of Results

The system integrates interactive visualizations to improve decision-making for both patients and healthcare providers.

1. Symptom-Disease Heatmap:
   - Displays relationships between common symptoms and diseases based on model predictions.
   - Helps doctors quickly identify patterns in symptom clusters.
2. Prediction Confidence Scores:
   - Each disease prediction is assigned a confidence score, indicating the model's certainty.
   - Provides transparency in AI-driven medical decision-making.
3. Historical Disease Trends:
   - Tracks disease prevalence based on user inputs over time.
   - Allows healthcare administrators to detect outbreaks and seasonal disease patterns.

E. Case Study: Real-World Testing

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

The system was tested using anonymized real-world symptom data from 50 patients. The predicted diseases were compared against actual diagnoses provided by medical professionals.

Fig. 3: Model Prediction vs Actual Diagnosis Table

Patient ID	Symptoms	Model Prediction	Doctor's Diagnosis	Correct Prediction
P001	Fever, Headache, Fatigue	Malaria	Malaria	True
P002	Chest Pain, Shortness of Breath	Pneumonia	Pneumonia	True
P003	Abdominal Pain, Nausea	Gastritis	Food Poisoning	False
P004	Skin Rash, Joint Pain	Dengue	Dengue	True
P005	High Fever, Cough	Typhoid	Typhoid	True

The success rate of predictions was 92%, confirming the model's effectiveness in disease identification. Incorrect predictions were analyzed, leading to improvements in feature selection and model tuning.

F. Limitations and Future Enhancements

1. Current Limitations:
   - The system relies on user-reported symptoms, which may be subjective.
   - It does not yet incorporate real-time clinical test results (e.g., blood reports, X-rays).
   - Some rare diseases with limited training data may be misclassified.
2. Future Enhancements:
   - Integration of wearable health device data for real-time monitoring.
   - Expanding the training dataset with electronic health records to improve accuracy.
   - Deployment as a mobile application to reach a wider audience.

The Random Forest-based disease prediction system achieves high accuracy and real-time usability for patient diagnosis. It offers an interactive PHP-based interface with real-time visualizations, ensuring that patients, doctors, and healthcare administrators can make informed decisions. Future improvements will focus on enhancing data integration and predictive capabilities to improve healthcare outcomes.

VI. CONCLUSION AND FUTURE WORK

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

A. Conclusion

The Integrated Health Management and Analytics System successfully combines machine learning with web-based healthcare analytics to provide an accurate and efficient disease prediction platform. By leveraging a Random Forest-based predictive model, the system achieves high accuracy, robustness, and reliability in diagnosing diseases based on user-reported symptoms.

The PHP-based frontend ensures user-friendly interaction, while the Python-integrated backend processes complex ML-based disease predictions. Through real-time visualization and analytics, the system enables patients, doctors, and healthcare administrators to make informed decisions regarding disease diagnosis and treatment.

Key contributions of this research include:

- Implementation of an advanced ML model (Random Forest) for disease prediction with an accuracy exceeding 95%.
- Seamless PHP-Python integration to enable AI-powered healthcare analytics.
- Interactive visualizations such as symptom-disease heatmaps, prediction confidence scores, and disease prevalence trends.
- Real-time processing and analysis, ensuring that disease predictions are generated within seconds.

The system's validation, based on performance metrics, real-world testing, and user feedback, confirms its effectiveness in assisting healthcare professionals and enhancing patient awareness.

B. Future Work

Although the system performs efficiently in symptom-based disease prediction, there are several areas for enhancement. Future developments will focus on expanding the model's capabilities and integrating additional healthcare data sources to improve accuracy and usability.

1. Incorporation of Clinical Test Data
   - Currently, predictions are made solely based on user-reported symptoms.
   - Future versions will integrate blood test results, X-ray reports, and vital signs (e.g., heart rate, oxygen levels) to improve diagnostic accuracy.
2. Integration with Wearable Devices
   - Real-time health monitoring will be enabled by connecting with smartwatches and wearable sensors to gather live health data (e.g., blood pressure, glucose levels).
   - This will allow for continuous monitoring and early disease detection.
3. Enhancement of the ML Model
   - Hybrid AI models (Deep Learning + Random Forest) will be explored to improve disease classification.
   - Automated feature engineering will be implemented to optimize symptom selection and improve predictive accuracy.
4. Expansion of the Disease Database
   - The current dataset is focused on common diseases. Future versions

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

will include rare diseases, genetic disorders, and region-specific illnesses to increase coverage.
- Collaboration with medical institutions to integrate electronic health records (EHR) will enhance the model's learning capabilities.

5. Mobile Application Development
   - A mobile-friendly application will be developed to increase accessibility.
   - Voice-based symptom input and chatbot-based diagnosis assistance will be implemented for better user engagement.
6. Cloud-Based Deployment for Scalability
   - Transitioning the system to a cloud-based platform will ensure scalability, security, and real-time global access.
   - This will allow multiple users (patients and doctors) to access the system simultaneously with minimal latency.
7. Collaboration with Healthcare Institutions
   - Partnership with hospitals, clinics, and research institutions will enable real-world testing and refinement.
   - Medical expert feedback will be used to fine-tune the AI model and improve diagnostic reliability.

C. Final Thoughts

The development of AI-driven healthcare solutions marks a significant advancement in medical diagnostics and patient care. The Integrated Health Management and Analytics System serves as a stepping stone toward AI-powered, personalized medicine. With further enhancements, the system has the potential to transform disease prediction and early diagnosis, thereby improving health outcomes on a large scale.

REFERENCES

[1] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324..,.

[2] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. Cambridge, UK: Cambridge University Press, 2014, doi: 10.1017/CBO9781107298019.

[3] A. Esteva et al., "A Guide to Deep Learning in Healthcare," *Nature Medicine*, vol. 25, no. 1, pp. 24–29, 2019,

[4] C. J. Kuo, M. G. Chang, and Y. H. Hsieh, "A Machine Learning-Based Disease Diagnosis System," *IEEE Access*, vol. 7, pp. 123419–123428, 2019.

[5] K. Rajpurkar et al., "Deep Learning for Chest X-ray Diagnosis," *IEEE Transactions on Medical Imaging*, vol. 37, no. 8, pp. 2104–2115, 2018..

[6] A. B. Smits and G. F. Hughes, "Feature Selection for High-Dimensional Biomedical Data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 4, pp. 1051–1062, 2021.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

[7] P. S. Yu, C. Faloutsos, and S. Papadimitriou, "Graph-Based Machine Learning for Disease Prediction," *Proceedings of the IEEE*, vol. 108, no. 11, pp. 2024–2043, 2020.

[8] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016

[9] R. Zhang, L. Cao, and J. Wang, "A Hybrid Random Forest Model for Healthcare Data Analysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 2983–2991, 2020

[10] Z. Li et al., "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," IEEE Access, Jun. 2021.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015

[12] A. R. Chowdhury et al., "Predictive Analytics in Healthcare: A Machine Learning Perspective," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 1, pp. 155–165, 2021

[13] H. R. Moghaddam et al., "A Comparative Analysis of Random Forest and Support Vector Machine for Disease Prediction," *IEEE Access*, vol. 8, pp. 202043–202056, 2020

[14] B. Y. Ribeiro, A. L. Silva, and M. H. Kimura, "Applying Random Forest for Disease Classification in Healthcare," *IEEE Latin America Transactions*, vol. 18, no. 4, pp. 648–655, 2020

[15] J. Brownlee, *Machine Learning Mastery With Python*. San Francisco, CA, USA: Machine Learning Mastery, 2016.

[16] M. H. He et al., "Efficient Feature Extraction Using Ensemble Learning for Medical Diagnosis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 12, pp. 5124–5135, 2020

[17] P. Gupta et al., "AI-Powered Diagnosis in Healthcare Using Hybrid Machine Learning Models," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 2, pp. 120–134, 2021.

[18] V. P. Chaurasia and S. Pal, "A Novel Approach for Early Detection of Chronic Diseases Using Random Forest," *IEEE Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 3, pp. 951–963, 2021.

[19] J. W. Lee, T. S. Kim, and K. R. Lee, "Machine Learning-Based Diagnosis and Prediction of Infectious Diseases," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 9, pp. 1–11, 2021.

[20] B. R. Pratt, M. J. Jones, and A. G. Smith, "Explainable AI in Healthcare: Improving Trust in Machine Learning Predictions," *IEEE Transactions on Artificial Intelligence*, vol. 3, no. 1, pp. 45–58, 2022.